

⑫ 公開特許公報(A) 平1-284898

⑤ Int. Cl.<sup>4</sup>  
G 10 L 3/00

識別記号 庁内整理番号  
H-8622-5D

④ 公開 平成1年(1989)11月16日

審査請求 未請求 請求項の数 1 (全4頁)

⑥ 発明の名称 音声合成装置

⑦ 特 願 昭63-115721

⑦ 出 願 昭63(1988)5月11日

⑦ 発 明 者 広 川 智 久 東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内

⑦ 出 願 人 日本電信電話株式会社 東京都千代田区内幸町1丁目1番6号

⑦ 代 理 人 弁理士 草 野 卓

明 細 書

1. 発明の名称

音声合成装置

2. 特許請求の範囲

- (1) 入力されたテキストに応じ、音声出力する  
音声合成装置において、

上記入力テキストを解析するテキスト解析部  
と、

そのテキスト解析部の出力を基に音声合成の  
ための韻律情報を生成する韻律情報生成部と、

音素など出力音声を組み立てる上で適切な単  
位毎に、原波形、発声された音韻環境、基本周  
波数パターン形状、継続時間情報、振幅情報など  
を記載した大量の波形情報を格納する波形辞書  
と、

上記テキスト解析部と上記韻律情報生成部か  
らの情報により、上記波形辞書より適切な波形  
を選択する波形選択部と、

所望の波形がない場合には最も選択条件に近  
い波形に対し使用目的に合致するよう変形を施

す波形変形処理部と、

所望の波形が全くない場合は、新たに波形を  
生成する波形生成部と、

これら波形選択部、波形変形処理部、及び波  
形生成部からの波形を接続する波形接続部とを  
備えることを特徴とする音声合成装置。

3. 発明の詳細な説明

「産業上の利用分野」

この発明は、テキストを入力しそのテキストに  
応じた任意の音声出力する音声合成装置、特に  
主に音韻系列と韻律情報とから音声を合成する規  
則合成装置に関するものである。

「従来の技術」

従来、任意の音声出力する規則合成装置では、  
音声合成方式にLPC(Linear Prediction  
Corder)方式を利用し、音声の結合単位にはCV  
やVCV、CVCなど音韻との対応や調音結合を  
考慮した単位を設定し、基本周波数パターンなどの  
韻律情報はアクセント形や呼吸段落内のモーラ数  
などから音韻情報とは独立に生成する方式を採っ

ているものが多い。しかしながらこれらの方法では、当然ながら音声合成時に分析時とは異なった基本周波数を持つ音源で駆動するため、LPCパラメータがあらわす声道スペクトルと音源スペクトルの不整合により、異常振幅やスペクトルQ値の低減が生じ、合成音品質劣化の原因となっている。これはLPC方式が声道パラメータと音源パラメータとの独立を仮定した分析合成方式であるにもかかわらず、実際にはこれら二つのパラメータは本来独立ではなく微妙に関係しあっているために起こる劣化で、規則合成にLPC分析合成手法を用いる根本的な問題と考えられる。

他に音声の特徴をホルマントで記述し、ホルマントの動きを規定することにより規則合成音を得る方式があるが、ホルマントの自動抽出が難しく、ホルマント遷移の記述も十分ではないため、LPCを用いる方式より品質が良くないのが現状である。

一方、このような問題を回避し、明瞭性の高い原波形に着目した方式もいくつか提案されている。しかしいずれも音素や音節単位に高々数種類の波

形選択部で選択され、所望の波形がない場合には最も選択条件に近い波形に対し使用目的に合致するように波形変形処理部で変形が施され、所望の波形が全くない場合は新たに波形が波形生成部で生成される。これら波形選択部、波形変形処理部及び波形生成部からの波形は波形接続部で接続される。

このようにこの発明によれば大量の波形を辞書として蓄積しておき、入力テキストに対し最も適した波形を選択して接続することにより出力音声を合成しているため、明瞭性が高く、しかも自然性の良い音声を得られる。

#### 「実施例」

第1図はこの発明の一実施例を示すブロック図である。すなわち端子1より音声に変換すべきテキストが入力されると、テキスト解析部2により係り受けや品詞解析などの形態素解析、および漢字かな変換、アクセント処理が行われ、音韻系列バッファ7、韻律情報生成部3に必要な情報が送出される。その情報としては音韻系列バッファ7

形を用意し、基本周波数や継続時間長の調整は、波形の打ち切りや繰り返し、間引き等を実施することにより対処している。従って合成音声は細かな制御は不可能であり、短かい音声をはりあわせた感じの音や、ブザーのような機械音的になってしまうという欠点を有していた。

この発明の目的は、テキスト合成に必要な規則合成において、自然性、明瞭性のともに高い合成音声の出力を可能とする音声合成装置を提供することにある。

#### 「課題を解決するための手段」

この発明によれば入力テキストはテキスト解析部で解析され、そのテキスト解析部の出力を基に音声合成のための韻律情報が韻律情報生成部で生成される。一方音素など出力音声を組み立てる上で適切な単位毎に、原波形、発声された音韻環境、基本周波数パターン形状、継続時間情報、振幅情報などを記載した大量の波形情報を格納する波形辞書が設けられ、テキスト解析部及び韻律情報生成部からの情報により波形辞書より適切な波形が波

に対しては音韻の区別を示す記号列、韻律情報生成部3に対しては呼気段落内モーラ数、アクセント形、発声スピードなどである。韻律情報生成部3はこれらの情報を基にピッチパターン、各音素毎の時間長パターン、および振幅パターンを規則により生成し、それぞれのバッファ4, 5, 6に書き込む。

波形選択部8は音韻系列バッファ7、ピッチパターンバッファ4、時間長バッファ5、振幅バッファ6を参照して、波形辞書9より最適な波形を選択する。波形辞書9は一例として、第2図に示すような構成をしており発声時の種々の情報とともに波形が格納されている。種々の情報とは、すなわち音韻種別、前後7程度の音韻環境、音素内の平均ピッチ、ピッチの形状を示すための1次直線で近似した場合の傾き、音素の継続時間長、波形中心部での数ピッチの始点・終点を示す時間長調整用情報、正規化した音素波形のRMS値(振幅)および実際の波形データである。波形辞書9の作成は、大量の発声データをもとにオフライン処理

で予め作成しておく。例えば男性アナウンサー名の発声による単語、文章など約数時間の音声データを12 KHzでA/D変換し、デジタルソナグラムの視察により音韻ラベリングを施す。この音声データに対し、ラベルの音韻境界前後20～30msの波形をディスプレイに表示し、カーソルで切り出すことで作成できる。切り出し位置は原則として波形の負から正への0切片とし、さらに音韻毎に例えば正のピークの手前の0切片で切り出すなど、ルールを定めておく。こうすることで接続点での不連続は避けられ、滑らかな連続波形が得られる。また音声データをLPC等で音声分析してピッチを抽出し、ピッチ形状や時間長などにおいて類似の音韻の統合化を行っておけば、切り出し波形数を低減でき、能率良く波形辞書の作成ができる。

波形選択部8の動作をさらに詳細に述べると、一例として第3図に示したようになる。まず検索音韻系列を設定する。検索音韻系列は該当する音韻を中心に置き、辞書中にある環境音韻の数での

次に振幅条件による選択を行う。この場合も時間長条件と同様、振幅バッファ6の振幅、および許容範囲から条件が設定される。波形候補が見つからない場合は、時間長と同様、最も条件に近い波形候補が選択され、波形変形処理部10において振幅調整処理を施す。こうして音韻環境や韻律条件により波形選択部8で選択された波形、波形生成部11で作られた波形、および波形変形処理部10で調整された波形は波形接続部12に送出され、順次結合されて音声波形として出力端子13に出力される。

波形生成部11では、例えばLPC技術を用いて任意のピッチを持つ波形を生成する。すなわち音韻対応にスペクトルを示すLPCパラメータを蓄積しておき、指定されたピッチによりパルス、または残差などを駆動し波形を生成する。ここでLPC技術を用いることは発明の目的と異なるが、この波形生成部11は波形が全くない場合の、いわば救済措置であり使用頻度は少ないと考えられる。

窓かけを行って入力音韻系列から切り出して設定する。波形辞書9を検索して波形候補が見つからない場合は順次検索音韻系列を両側から削除していきながら検索を行う。検索音韻系列が該当する音韻のみとなっても、波形候補が見つからない場合、波形生成部11において所望のピッチ波形の生成を行う。次に合成音声の自然性に最も大きな影響を及ぼすと考えられるピッチパターンを考慮し、選択すべき音素のピッチ条件を設定する。これはピッチパターンバッファ4を参照して、平均ピッチ、ピッチの形状より決定する。許容範囲は実験値より決定すべきであるが、およそ所望値の5%以内ならば自然性は保たれると考えられる。波形候補が見つかった場合は、それらに対し時間長条件による選択を行う。時間長条件は、時間長バッファ5の時間長と、ピッチ条件と同様に、実験値より決まる許容範囲とから設定される。時間長条件に合う波形候補がない場合、最も条件に近い波形候補が選択され、波形変形処理部10において時間長調整処理を施す。波形候補が見つかった場合は、

また波形変形処理部10では、時間長調整処理、振幅調整処理を行っている。以下にそれらの処理について説明する。

時間長調整処理は、当該音韻が無声音と有声音で処理が異なる。無声音の場合、破裂音であれば無音区間を伸縮する事で対処し、摩擦音であれば中心部から前後に向かって所望の時間長になるよう、波形の切断、または繰り返し使用を行う。有声音の場合は、波形中心部でのピッチ位置を3ピッチ程度辞書中に蓄えておき、波形データの方が長い場合はそれらの間引き、波形データの方が短い場合はそれらの繰り返し使用を行う。

振幅調整処理は、音素毎に定められた振幅値を振幅バッファより参照して、選択または生成された波形のRMS値との比率により振幅値を線形に調整する。

「発明の効果」

以上述べたようにこの発明によれば、大量の波形を辞書として蓄積しておき、入力テキストに対し最も適した波形を選択し、接続することで出力

音声を合成しているため、明瞭性が高く、しかも自然性も良い音声を提供できる。

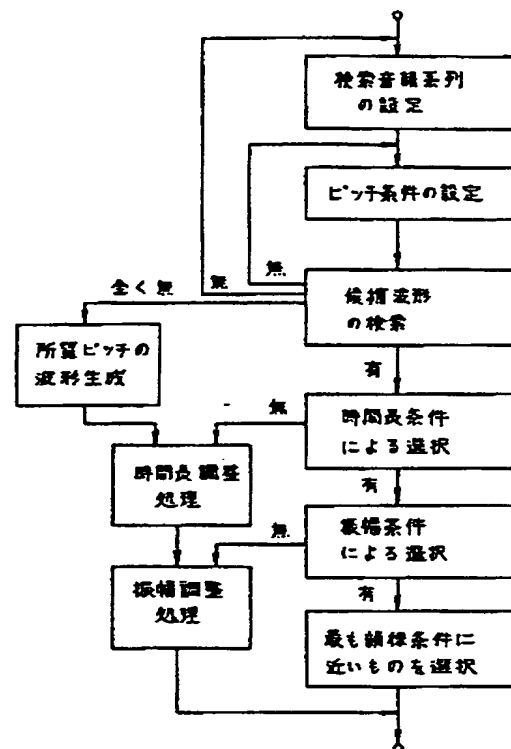
#### 4. 図面の簡単な説明

第1図はこの発明による音声合成装置の実施例を示すブロック図、第2図は波形辞書9の一構成例を示す図、第3図は波形辞書から最も適切な波形を選択する方法を示すフロー図である。

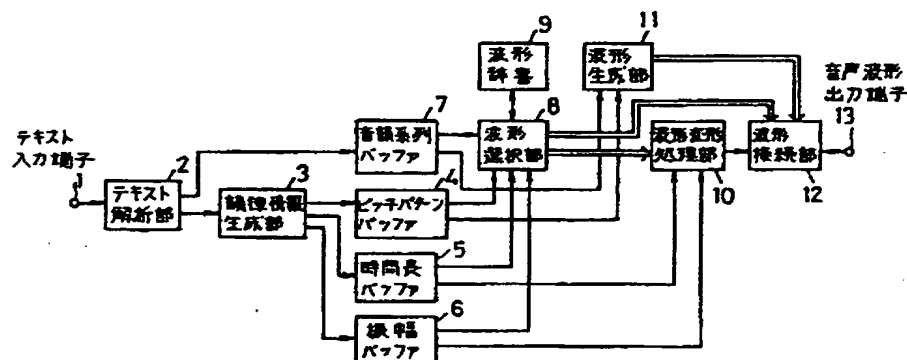
特許出願人 日本電信電話株式会社

代理人 草野 卓

カ 3 図



カ 1 図



カ 2 図

音韻	環境	平均 ビッチ	ビッチの 傾き	時間長	調整ポイント	振幅	波形
ɑ	sekaine	100	+0.15	120ms	64,136,230	150	+++++
ɑ	ouwa	120	-0.10	150ms	39,159,297	1000	~~~~~
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	toshiwa	*	*	200ms	*	120	~~~~~
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮